BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(19) JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **11045097 A**

(43) Date of publication of application: **16 . 02 . 99**

(51) Int. Cl **G10L 3/00**
G10L 5/06

(21) Application number: **09201685**

(22) Date of filing: **28 . 07 . 97**

(71) Applicant: **NEC CORP**

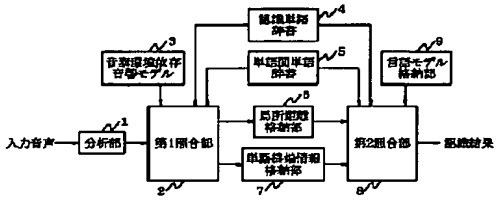
(72) Inventor: **ISOTANI RYOSUKE**

(54) **CONTINUOUS VOICE RECOGNIZION SYSTEM**

(57) Abstract:

PROBLEM TO BE SOLVED: To perform large vocabulary continuous voice recognizing at high speed using a phoneme environment depending acoustic model.

SOLUTION: A recognition word dictionary 4 describes an acoustic model group decided without depending on words of before and after about each word of a recognition object vocabulary as a recognition word. An inter-word dictionary 5 describes an acoustic model group used depending on words of before and after in a word border as inter-word. A first collating section 2 collates a time series of a feature parameter obtained by analyzing an input voice with the recognition word and the inter-word, and a word corresponding to a score when it is assumed that the group reaches a terminal of a word in each time of a time series of the feature parameter is outputted as word terminal information 7. A second collating section 8 refers to this word terminal information, collates again a time series of the feature parameter with the recognition word and the inter-word, and a word corresponding to a prescribed score is outputted with a mode decided by a system.



(19)日本国特許庁（ＪＰ）　(12) 公開特許公報（A）　(11)特許出願公開番号

特開平11－45097

(43)公開日　平成11年(1999)２月16日

(51)Int.Cl.[6]		識別記号	FI		
G10L	3/00	531	G10L	3/00	531D
	5/06			5/06	D

審査請求　有　　請求項の数5　OL　（全 6 頁）

(21)出願番号　特願平9－201685

(22)出願日　平成９年(1997)７月28日

(71)出願人　000004237
日本電気株式会社
東京都港区芝五丁目７番１号

(72)発明者　磯谷　亮輔
東京都港区芝五丁目７番１号　日本電気株式会社内

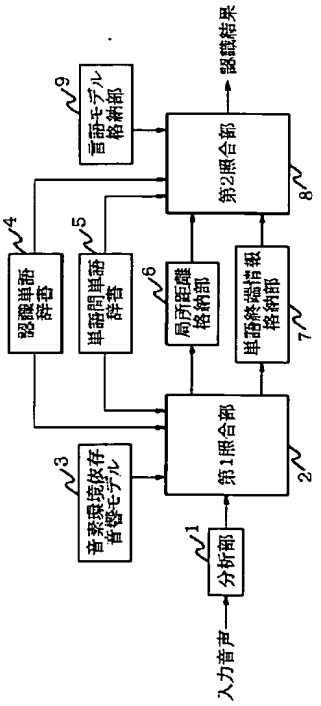
(74)代理人　弁理士　京本　直樹　（外２名）

(54)【発明の名称】　連続音声認識方式

(57)【要約】

【課題】　音素環境依存音響モデルを用いた大語彙連続音声認識を高速に行う。

【解決手段】　認識単語辞書４は認識対象語彙中の各単語について前後の単語に依存せずに決まる音響モデル系列を認識単語として記述する。単語間単語辞書５は単語境界において前後の単語に依存して用いられる音響モデル系列を単語間単語として記述する。第１の照合部２は入力音声を分析して得られる特徴パラメータの時系列を前記認識単語および単語間単語と照合して該特徴パラメータの時系列の各時刻において単語の終端に到達したと仮説した場合のスコアと対応する単語を単語終端情報７として出力する。第２の照合部８はこの単語終端情報を参照して再度特徴パラメータの時系列を前記認識単語および単語間単語と照合し予め定められたスコアに対応する単語をシステムで定められた形態で出力する。



1

【特許請求の範囲】

【請求項１】　音素環境依存音響モデルを用いる連続音声認識方式において、認識対象語彙中の各単語について前後の単語に依存せずに決まる音響モデル系列を認識単語として記述した認識単語辞書と、単語境界において前後の単語に依存して用いられる音響モデル系列を単語間単語として記述した単語間単語辞書と、入力音声を分析して特徴パラメータの時系列を得る分析部と、前記特徴パラメータの時系列を前記認識単語および前記単語間単語と照合して該特徴パラメータの時系列の各時刻にて単語の終端に到達した仮説のスコアと対応する単語を単語終端情報として出力する第１の照合部と、前記単語終端情報を参照して再度前記特徴パラメータの時系列を前記認識単語および前記単語間単語と照合し照合スコアに基づいて単語列の候補をシステムで定められた形態で出力する第２の照合部を有することを特徴とする連続音声認識方式。

【請求項２】　前記第１の照合部が前記各時刻において前記スコアがあらかじめ定めた基準を満たさない仮説については照合を打ち切り、前記単語終端情報として各時刻において少なくとも単語終端に到達した仮説に対応する認識単語および単語間単語を出力することを特徴とする請求項１記載の音声認識方式。

【請求項３】　前記第１の照合部で各時刻ごとに計算される局所距離を格納する局所距離格納部を備え、前記第２の照合部が前記特徴パラメータの時系列に換えて前記局所距離格納部に格納された局所距離を用いて前記認識単語および前記単語間単語と照合することを特徴とする請求項１及び２記載の音声認識方式。

【請求項４】　前記第２の照合部が前記入力音声の最終時刻から逆向きに前記特徴パラメータの時系列を前記認識単語および前記単語間単語と照合することを特徴とする請求項１記載の音声認識方式。

【請求項５】　前記第２の照合部が前記入力音声の最終時刻から逆向きに前記局所距離を用いて前記認識単語および前記単語間単語と照合することを特徴とする請求項３記載の音声認識方式。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は連続音声認識方式に関し、特に、音素環境依存音響モデルを用いた大語彙連続音声認識のサーチ方式に関するものである。

【０００２】

【従来の技術】従来、大語彙連続音声認識を高速に行う方法として、まず第一ステップ（ｆｏｒｗａｒｄ　ｐａｓｓ）でフレーム同期のビタビビームサーチを行って各フレームでアクティブ（ビーム内に残っている）であった単語の終端の情報を保存し、つぎに第二ステップ（ｂａｃｋｗａｒｄ　ｐａｓｓ）で、第一ステップで保存した単語終端情報を利用して、入力の終端から逆向きにビタビビームサーチを行う方法が知られている（“Ｔｈｅ　Ｆｏｒｗａｒｄ－Ｂａｃｋｗａｒｄ　Ｓｅａｒｃｈ　Ａｌｇｏｒｉｔｈｍ”，Ｐｒｏｃ．ＩＣＡＳＳＰ－９１，　ｐｐ．６９７－７００（以下、引用文献１と称す））。第二ステップでは、各フレームにおいて単語始端から先行する単語へ仮説を展開する際に、第一ステップで終端がアクティブでなかった単語や、あるいはスコアの悪かった単語については探索を行わないようにすることにより、大幅に計算量を削減できる。

2

【０００３】一般に、第一ステップでは単語ごとに仮説を保持するなど精度の粗い探索を行い、第二ステップでは、たとえばＮ－ｂｅｓｔ（上位Ｎ候補）を求めるために連続する２単語の組ごとに仮説を保持するなど、第一ステップより精度の細かい探索を行う。これにより、最初から精度の細かい探索を行う場合に比べ、全体の計算量を大幅に削減することが可能となる。

【０００４】一方、一般に音声認識において、音素等を単位とした音響モデルを用いる場合、前後の音素環境（コンテキスト）に依存したモデルが有効であることが知られている。たとえば、前後一つずつの音素に依存した音素モデルであるトライフォンモデルは広く使われており、また、前後の一方だけの音素や、前後二つ以上の音素に依存した音素モデルを用いることもある。連続音声認識の場合、単語内だけでなく単語境界にも音素環境依存の音響モデルを用いる方が精度がよいことが知られているが、単語の始終端に用いる音響モデルが前後に接続する単語に依存するため、環境に依存しない音響モデルを用いる場合に比べ、処理量が大幅に増えてしまう。この状況を簡略化した例を図７を用いて説明する。

【０００５】図７は「ここ」「そこ」「から」「まで」の４単語が自由な順序でつながる連続音声を認識する場合で、音響モデルとして後続音素に依存する音素モデルを用いる例である。たとえば、後続音素が／ｏ／である音素／ｋ／の音響モデルを“ｋ（ｏ）”などと表している。図中、黒丸は単語終端を表し、前記第一ステップで、各フレームごとにアクティブであったものについて、その情報を保存する。

【０００６】各単語の終端について、後続する単語として４単語の可能性があり、その最初の音素としては／ｋ／，／ｓ／，／ｍ／の３通りの可能性があるので、各単語の終端に用いる音響モデルはそれぞれに応じて音素環境依存モデルを用意する必要がある。したがって、全体として照合すべき音響モデルの数は２４となる。大語彙の場合は、単語終端における後続音素の種類数が多くなるため、単語境界での音響モデル数の増大はさらに大きく、処理量が大幅に増大する。実際にはビームサーチを行っているため、常にすべてのモデルと照合するわけではないが、全体の探索空間が広くなるため精度を確保するためにはビーム幅を大きくとらなければならず、結果的に処理量が増大することになる。

【０００７】この問題に対し、単語内には音素環境依存の音響モデルを用い、単語境界では環境に依存しない音響モデルを使用する連続音声認識方式が、特開平５－２２４６９２に開示されている。これにより、単語間での処理量の増大を抑えることができる。これを前述の４単語の例に適用すると、図８のようになる。ここで、たとえば、“ｏ（＊）”は後続音素によらない／ｏ／の音響モデルを表す。全体として照合すべき音響モデルの数は１６に削減される。大語彙の場合は、さらに削減効果は大きい。

【０００８】

【発明が解決しようとする課題】上述した従来の連続音声認識方式は、精度を向上させるために音素環境依存モデルを使用すると、単語境界での処理量が増大するという問題点がある。これに対し、単語内には音素環境依存の音響モデルを用い、単語境界では環境に依存しない音響モデルを使用すると、単語境界での処理量の増大を抑えることができるが、一方で単語境界に用いる音響モデルの精度が低いために、とくに大語彙の連続音声認識では性能の低下を生じるおそれがある。

【０００９】本発明の目的は、単語境界にも音素環境依存の音響モデルを用いて精度を確保しつつ、大語彙の場合でも単語境界での処理量の増大を抑えることのできる連続音声認識方法を提供することにある。

【００１０】

【課題を解決するための手段】本第１の発明の連続音声認識方式は、音素環境依存音響モデルを用いる連続音声認識方式において、認識対象語彙中の各単語について前後の単語に依存せずに決まる音響モデル系列を認識単語として記述した認識単語辞書と、単語境界において前後の単語に依存して用いられる音響モデル系列を単語間単語として記述した単語間単語辞書と、入力音声を分析して特徴パラメータの時系列を得る分析部と、前記特徴パラメータの時系列を前記認識単語および前記単語間単語と照合して該特徴パラメータの時系列の各時刻にて単語の終端に到達した仮説のスコアと対応する単語を単語終端情報として出力する第１の照合部と、前記単語終端情報を参照して再度前記特徴パラメータの時系列を前記認識単語および前記単語間単語と照合し照合スコアに基づいて単語列の候補をシステムで定められた形態で出力する第２の照合部を有することを特徴とする。

【００１１】また、第２の発明は、第１の発明における前記第１の照合部が前記各時刻において前記スコアがあらかじめ定めた基準を満たさない仮説については照合を打ち切り、前記単語終端情報として各時刻において少なくとも単語終端に到達した仮説に対応する認識単語および単語間単語を出力することを特徴とする。

【００１２】また、第３の発明は、第１の発明および第２の発明における前記第１の照合部で各時刻ごとに計算される局所距離を格納する局所距離格納部を備え、前記

第２の照合部が前記特徴パラメータの時系列に換えて前記局所距離格納部に格納された局所距離を用いて前記認識単語および前記単語間単語と照合することを特徴とする。

【００１３】また、第４の発明は、第１の発明における前記第２の照合部が前記入力音声の最終時刻から逆向きに前記特徴パラメータの時系列を前記認識単語および前記単語間単語と照合することを特徴とする。

【００１４】また、第５の発明は、第３の発明における前記第２の照合部が前記入力音声の最終時刻から逆向きに前記局所距離を用いて前記認識単語および前記単語間単語と照合することを特徴とする。

【００１５】

【発明の実施の形態】次に、本発明の実施の一形態について図面を参照して説明する。

【００１６】図１は、本発明の実施の一形態を示すブロック図である。

【００１７】図１において、入力音声は分析部１で特徴パラメータの系列に変換され、第１照合部２に入力される。第１照合部２では、認識単語辞書４、単語間単語辞書５および音素環境依存音響モデル３を参照して、引用文献１（従来技術を参照）におけるｆｏｒｗａｒｄ　ｐａｓｓと同様にフレーム同期ビタビビームサーチにより特徴パラメータ系列の照合を行う。音響モデルとしては隠れマルコフモデル（ＨＭＭ）を用いる。

【００１８】認識単語辞書４には、認識対象語彙の各単語について、その単語を構成する音響モデル系列のうち先行あるいは後続する単語に依存しない部分、および単語の表記が記述されている。単語間単語辞書５には、単語が隣接する場合に、前の単語を構成する音響モデルのうち後の単語に依存する部分と、後の単語を構成する音響モデルのうち前の単語に依存する部分を連結した音響モデル系列が、前後の単語の条件とともに記述されている。照合の際には、認識単語および単語間単語から図２に示すような単語のネットワークを構成し、これをさらにＨＭＭの状態の系列に展開する。各単語の始端から照合を行い、各フレームごとに、認識単語および単語間単語のうち最終状態がアクティブなものを、そのフレームまでの累積スコアとともに単語終端情報格納部７に保存する。また、各フレームごとのＨＭＭの各状態の出力確率を、局所距離格納部６に保存する。

【００１９】第１照合部での照合が終わると、続いて第２照合部８で、再び認識単語辞書４と単語間単語辞書５を参照して、引用文献１（従来技術を参照）におけるｂａｃｋｗａｒｄ　ｐａｓｓと同様に入力音声の終端から逆向きにビタビビームサーチにより照合を行い、最終的な認識結果を出力する。出力する認識結果は、最も簡単にはスコアのもっともよい１つの単語列であるが、この他スコアのよい上位複数の単語列としたり、単語のネットワーク（単語グラフ）としたりすることが出来る。そ

の際、言語モデル格納部９に記述された単語間の接続情報などを制約として用いる。

【００２０】言語モデルとしては、たとえば単語バイグラムモデルを用いることができる。一般には第一照合部と第二照合部で異なる音響モデルや異なる特徴パラメータを用いることもできるが、同じ音響モデルと特徴パラメータを用いる場合には、第一照合部で計算し局所距離格納部６に保存した局所距離の値を用いることができる。第二照合部では、アクティブな単語（単語間単語を含む）の始端に対して先行する単語のうち第一照合部で終端がアクティブであった単語についてのみ仮説を展開する。また、第一照合部で終端がアクティブであっても、その累積スコアと、第二照合部でのその時点までの累積スコアの値に応じて、見込みの小さい単語については仮説を展開しないようにすることもできる。

【００２１】以下、認識単語辞書４と単語間単語辞書５について、従来技術の説明に用いた簡略化した例に基づいて、図２を参照して説明する。この例では、音響モデルとして後続の音素に依存する音素モデルを使用しており、辞書は「ここ」「そこ」「から」「まで」の４単語からなる。たとえば、「そこ」は、／ｓ／，／ｏ／，／ｋ／，／ｏ／の４音素からなり、前の３つの音素に対応する音響モデルはそれぞれ“ｓ（ｏ）”，“ｏ（ｋ）”，“ｋ（ｏ）”と決まるが、最後の／ｏ／については、後続に「から」がくる場合は“ｏ（ｋ）”、「まで」がくる場合は“ｏ（ｍ）”となるなど、一意に決まらない。そこで、認識単語辞書では、単語「そこ」の音響モデル系列として“ｓ（ｏ）”，“ｏ（ｋ）”，“ｋ（ｏ）”のみを記述し、“ｏ（ｋ）”，“ｏ（ｍ）”などは可能なすべての種類を用意し、単語間単語として独立に扱う。単語間単語は、複数の単語の組合せで共有される。たとえば、“ｏ（ｋ）”，“ｏ（ｍ）”などは、先行単語が「ここ」など最後の音素が／ｏ／である単語に共通に用いることができる。また、“ｏ（ｋ）”は先頭の音素が／ｋ／である単語、“ｏ（ｍ）”は先頭の音素が／ｍ／である単語にのみ接続しうる。図中、黒丸は単語終端を表す。

【００２２】図２の例に対応した認識単語辞書４および単語間単語辞書５の構成例を、それぞれ図３、図４に示す。単語辞書には、認識対象語彙中の各単語に対し、表記、その単語を構成する音響モデル系列の情報に加え、始端、終端のカテゴリが格納されている。単語間辞書には、先行単語の終端カテゴリと後続単語の始端カテゴリの組合せに対し、それらの単語間に用いられる音響モデル系列の情報が記述されている。

【００２３】図２では全体として照合すべき音響モデルの数は２１で、図７に示した従来例の場合の２４に比べ削減されている。この例は　４単語からなる簡略化した例であるため差はそれほど大きくないが、大語彙の場合には後続音素環境の種類がふえるため、削減効果は大き

い。図８に示した従来例に比べると照合すべき音響モデルの数は増えているが、単語間でも図２の例と同様に音素環境依存音響モデルを使用しているために、精度の高い照合が可能となっている。

【００２４】音素環境としては、音素そのものでなく、いくつかの音素を一まとめにした音素クラスを用いることもできる。また、後続音素にのみ依存する音響モデルのかわりに、前後の音素に依存する音響モデルを用いることもできる。その場合は、認識単語辞書には最初の音素と最後の音素を除いた音響モデルを記述し、後続音素に依存する単語終端の音素と先行音素に依存する単語始端の音素に対応する音響モデルは、お互いに接続しうるものを組にして単語間単語辞書に記述しておけばよい。前後の音素に依存する音響モデルを用いる場合の認識単語辞書および単語間単語辞書の記述例を、それぞれ図５，図６に示す。図で、たとえば“（ｓ）ｏ（ｋ）”は、先行音素が／ｓ／で後続音素が／ｋ／である音素／ｏ／の音響モデルを表す。

【００２５】図２の例では、各単語の音響モデル系列を独立に扱っているが、一般に大語彙の場合は先頭部分が共通な単語が多く、全単語の音響モデル系列を先頭を共通化した木構造の形で表現することも可能である。

【００２６】

【発明の効果】以上説明したように、本発明によれば大語彙連続音声認識を、音素環境依存音響モデルを用いて高精度に、しかも従来の音素環境依存音響モデルを用いた場合よりも処理量を削減して行うことができる効果がある。

【図面の簡単な説明】

【図１】本発明の実施の一形態を示すブロック図である。

【図２】図１の第１照合部で構成される単語のネットワークの例を示す図である。

【図３】本発明の実施例における、認識単語辞書の例を示す図である。

【図４】本発明の実施例における、単語間単語辞書の例を示す図である。

【図５】本発明の実施例における、認識単語辞書の別の例を示す図である。

【図６】本発明の実施例における、単語間単語辞書の別の例を示す図である。

【図７】従来例における、単語終端の音響モデルの扱い示す図である。

【図８】従来例における、別の単語終端の音響モデルの扱いを示す図である。

【符号の説明】

１　分析部
２　第１照合部
３　音素環境依存音響モデル
４　認識単語辞書

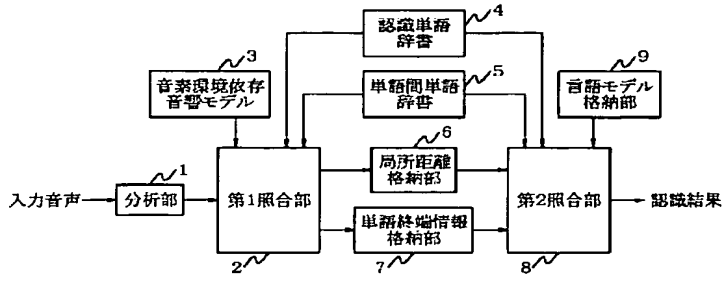
7

5　単語間単語辞書
6　局所距離格納部
7　単語終端情報格納部

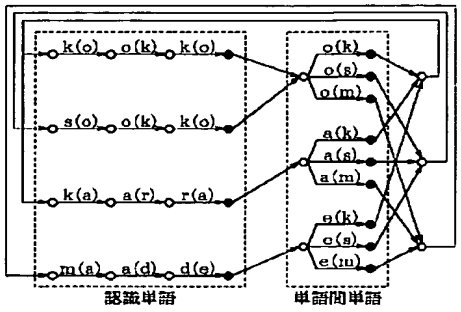
8

8　第２照合部
9　言語モデル格納部

【図１】



【図２】



【図３】

番号	表記	音響モデル列	始端	終端
1	ここ	k(o) o(k) k(o)	k	o
2	そこ	s(o) o(k) k(o)	s	o
3	から	k(a) a(r) r(a)	k	a
4	まで	m(a) a(d) d(e)	m	e

【図４】

番号	先行単語終端	後続単語始端	音響モデル列
1	o	k	o(k)
2	o	s	o(s)
3	o	m	o(m)
4	a	k	a(k)
5	a	s	a(s)
6	a	m	a(m)
7	e	k	e(k)
8	e	s	e(s)
9	e	m	e(m)

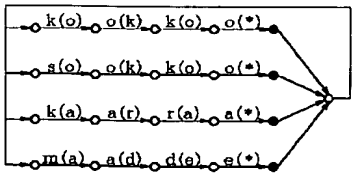
【図５】

番号	表記	音響モデル列	始端	終端
1	ここ	(k)o(k) (o)k(o)	ko	ko
2	そこ	(s)o(k) (o)k(o)	so	ko
3	から	(k)a(r) (a)r(a)	ka	ra
4	まで	(m)a(d) (a)d(e)	ma	de

【図６】

番号	先行単語終端	後続単語始端	音響モデル列
1	ko	ko	(k)o(k) (o)k(o)
2	ko	so	(k)o(s) (o)s(o)
3	ko	ka	(k)o(k) (o)k(a)
4	ko	ma	(k)o(k) (o)m(a)
5	ra	ko	(r)a(k) (o)k(o)
6	ra	so	(r)a(s) (o)s(o)
7	ra	ka	(r)a(k) (o)k(a)
8	ra	ma	(r)a(k) (o)m(a)
9	de	ko	(d)e(k) (o)k(o)
10	de	so	(d)e(s) (o)s(o)
11	de	ka	(d)e(k) (o)k(a)
12	de	ma	(d)e(k) (o)m(a)

【図８】



【図７】